Artificial Intelligence & AI Programming

3 main course components :

1. **Search Methods**

2. **Logic & Resolution**

3. **Uncertainty Reasoning**

   - Probabilistic reasoning
   - Bayes’ theorem
   - Belief networks
   - Dempster-Shafer theory
   - Fuzzy inference

Lecture 9: Introduction to Bayesian Methods

Topics: / key words

- Why probability?
- Review probability
- Bayes' Theorem
- Normalisation

Lecture 9: Introduction to Bayesian Methods

Diagnosis

Logic rules or …..??

Lecture 9: Introduction to Bayesian Methods

Diagnosis

Logic rules or …..??

$\forall x \exists y\ [\text{Symptom}(x) \Rightarrow \text{Disease}(y)]$

$[\text{Symptom}(\text{toothache}) \Rightarrow \text{Disease}(\text{Cavity})]$

$[\text{Symptom}(\text{toothache}) \Rightarrow \text{Disease}(\text{Cavity}) \vee$

$\text{Disease}(\text{impacted wisdom tooth}) \vee$

$\text{Disease}(\text{gingivitis}) \vee \ldots\ldots \text{Disease}(\ ) \ldots\ldots]$

Logic fails due to…..

- Laziness
- Theoretical/practical ignorance

Basic Probability - 1

Simple rules and axioms of classical probability

Uncertain evidence arises because of …..

- **INCOMPLETE** - missing/unavailable data
- **INEXACT** - poorly measured data
- **IMPRECISE** - random statistics

Probability usually defined as

- **Prob( ) =** $\dfrac{\textbf{\# desired outcomes}}{\textbf{total \# outcomes}}$

e.g. Find probability of drawing a heart from pack of cards

 # hearts = 13

 # cards = 52

P(heart) = $\dfrac{13}{52} = \dfrac{1}{4}$

Basic Probability - 2

Simple rules and axioms of classical probability

Previous example utilises the axioms ….

- **All probabilities lie between 0 and 1**

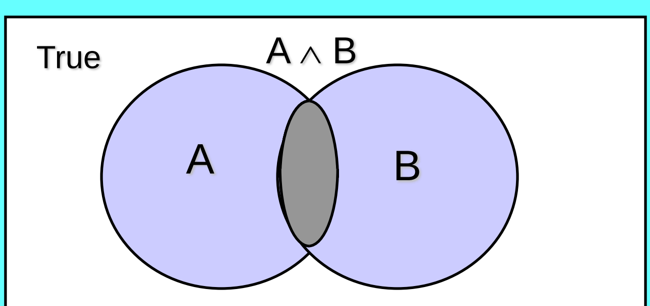
$$0 \leq P() \leq 1$$

- **True (i.e. valid) propositions have P() = 1 and false (i.e. unsatisfiable) propositions, P() = 0**

$$P(True) = 1 \qquad P(False) = 0$$

- **Probability of a disjunction is given by**

$$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$$



Basic Probability - 3

Simple rules and axioms of classical probability

Further properties can be derived from the previous axioms, for example

$$P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$$

$$P(True) = P(A) + P(\neg A) - P(False)$$

$$1 = P(A) + P(\neg A)$$

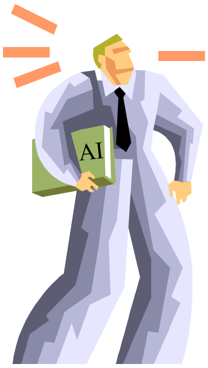
$$P(\neg A) = 1 - P(A)$$

Basic Probability - 4

Conditional Probability

unconditional

prior P(H)



conditional

posterior P(H |E)

H

hypothesis

E

symptom

Conditional probability of H occurring, given evidence E is

- $$P(H|E) = \frac{P(H \wedge E)}{P(E)}$$, provided P(E)>0

i.e. $P(H \wedge E) = P(H|E)\,P(E)$ $\boxed{\text{PRODUCT RULE}}$

$P(E \wedge H) = P(E|H)\,P(H)$

The Joint Probability Distribution

$\mathbf{P}(X_1 \ldots\ldots X_n)$ is a 1D vector of probabilities for the possible values of the variable X_i.

The joint probability distribution assigns probabilities to all propositions in the domain.

Then the joint is an *N*-D table with a value in every cell giving the probability of that specific state occurring.

	Toothache	¬Toothache	
Cavity	0.04	$\boxed{0.06}$	P(C)
¬Cavity	0.01	0.89	1-P(C)
	P(T)	1-P(T)	

Analysis between Sleep Stages and Sleep Poses:

$P(N2 \wedge Left) = 281/1158$

Sleep Stages (SS)

Sleep Poses (SP)	N1	N2	N3	REM	Awake	Total # of each SP
Empty Bed	0	0	0	0	24	24 P(EB)
Prone	9	0	0	0	27	36 P(Prone)
Right	0	47	60	0	7	**114** P(Right)
Left	63	281	55	98	82	**579** P(Left)
Supine	12	259	80	0	54	**405** P(Supine)
Total # of each SS	**84** P(N1)	**587** P(N2)	**195** P(N3)	**98** P(REM)	**194** P(Aw)	**= 1158** = 100%

$P(Left) = P(579/1158)$
$= \sum P(Left \wedge SS)$

$P(N2) = 587/1158$

Conditional probability of H occurring, given evidence E is

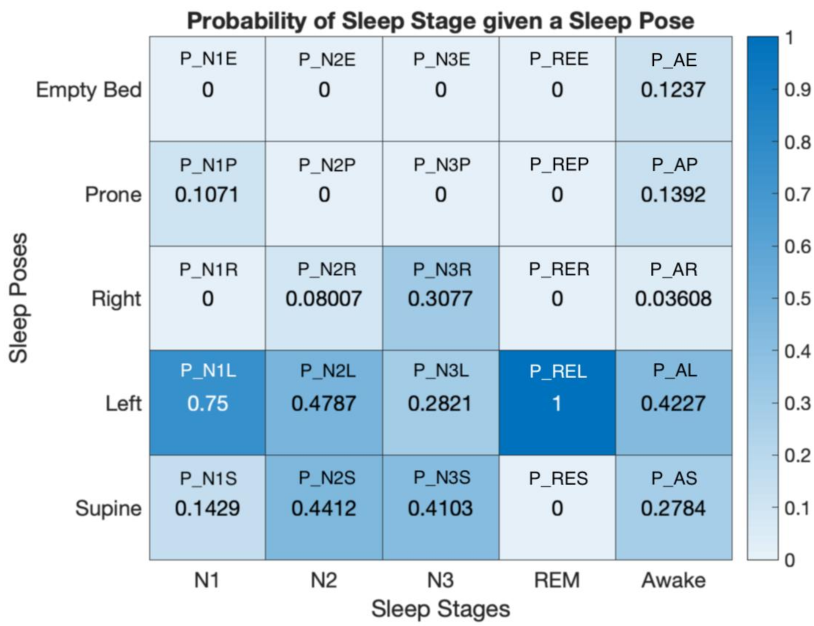
- $P(H|E) = \dfrac{P(H \wedge E)}{P(E)}$, provided $P(E)>0$

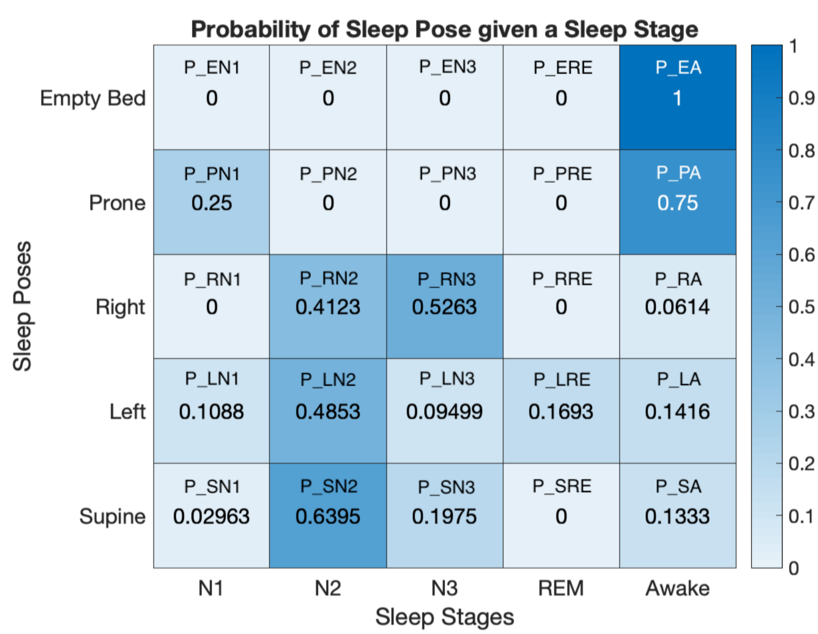
i.e. $P(H \wedge E) = P(H|E)\,P(E)$
$P(E \wedge H) = P(E|H)\,P(H)$

PRODUCT RULE

From S Mahvash et al, University of Surrey, 2020
Published in IEEE Trans Biomed Eng

Analysis between Sleep Stages and Sleep Poses:





Lecture 9: Introduction to Bayesian Methods

End of segment 1

Bayes' Rule

After Reverend Thomas Bayes (1702-1761).

Used as a cornerstone for AI reasoning systems since 1960s, especially in medical diagnosis (e.g. MYCIN).

Recall 2 forms of product rule

- **i.e.** $P(H \wedge E) = P(H \mid E)\, P(E)$

  $P(E \wedge H) = P(E \mid H)\, P(H)$

Equate 2 RHS's and ÷ P(E) …

$$P(H \mid E) = \frac{P(E \mid H)\, P(H)}{P(E)}$$

Bayes' Rule

Example: Meningitis Problem

Consider calculating the probability of meningitis, given the symptoms of a stiff neck…..

Background information :

- Half of all meningitis patients have a stiff neck;
- Incidence of meningitis = 1 / 50,000
- Incidence of a stiff neck = 1/20.

What is the probability of meningitis, given a stiff neck?

Let S be the proposition that the patient has a stiff neck

Let M be the proposition that the patient has meningitis

$P(S \mid M) = 0.5$, $P(M) = 1/50k$, $P(S) = 1/20$

$$P(M \mid S) = \frac{P(S|M)\, P(M)}{P(S)} \qquad \text{(Bayes' Rule)}$$

$$= \frac{0.5 \quad 1/50k}{1/20} = \underline{\mathbf{0.0002}}$$

Normalisation

Sometimes, exact knowledge of the prior (e.g. P(S) in last example) may be unknown or difficult to evaluate accurately.

Use Normalisation :

First write:
$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E)} \qquad \ldots 1$$

Also:
$$P(\neg H|E) = \frac{P(E|\neg H)\,P(\neg H)}{P(E)} \qquad \ldots 2$$

Adding 1 and 2 gives:
$$P(\neg H|E) + P(H|E) = \frac{P(E|H)\,P(H) + P(E|\neg H)\,P(\neg H)}{P(E)} \qquad \ldots 3$$

But: $P(\neg H|E) + P(H|E) = 1$

So using RHS of equation 3…
$$P(E) = P(E|H)\,P(H) + P(E|\neg H)\,P(\neg H)$$

Substitute back into equation 1 (Bayes' Rule):
$$P(H|E) = \frac{P(E|H)\,P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)} \qquad \ldots 4$$

$= P(E \wedge H) + P(E \wedge \neg H)$

Normalisation

Thus determining the extra conditional probability $P(E|\neg H)$ allows us to avoid evaluating the prior P(E) directly. In general we can write:

$$\mathbf{P}(H|E) \quad = \quad \alpha\, \mathbf{P}(E|H)\, \mathbf{P}(H) \qquad \ldots 5$$

where α is a normalisation constant needed to ensure that entries in the joint $\mathbf{P}(H|E)$ sum to 1.

Bayesian Updating Rule

When calculating probabilities, we are often faced with the need to update our estimates in the light of new evidence.

Bayesian updating rule :

$$P(H|E_1,E_2) = \alpha\, P(H)\, P(E_1|H)\, P(E_2|H)$$

where α is a normalisation constant needed to ensure that entries in the joint $\mathbf{P}(H|E_1,E_2)$ sum to 1.

Assumption:
This relationship only holds true when conditional independence is true.
i.e. We assume that, in this particular case, the conditional probability $P(E_1|H \wedge E_2)$ does not depend on E_2 . Similarly, we assume that the conditional probability $P(E_2|H \wedge E_1)$ does not depend on E_1.

Formally we can write:

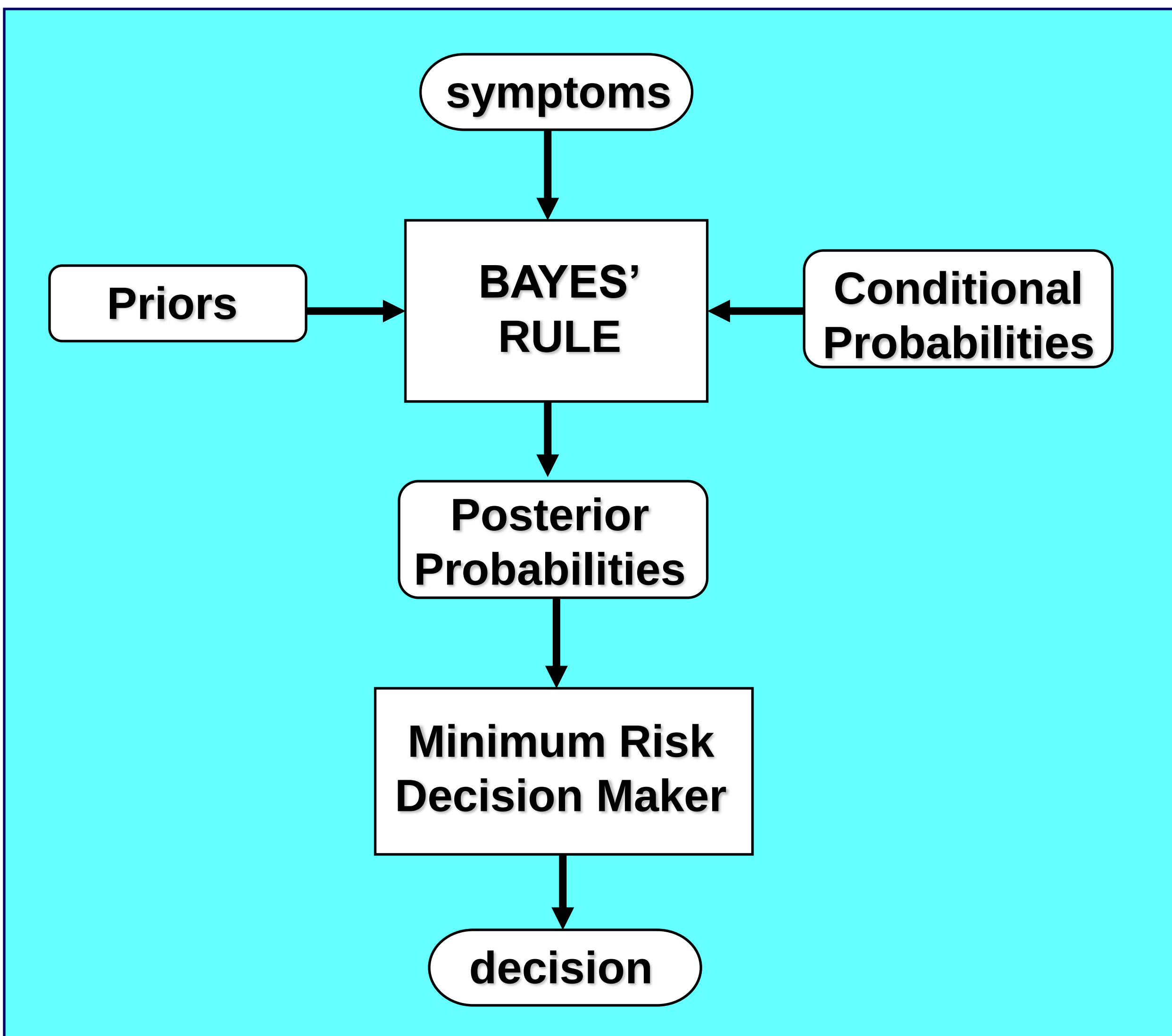
$$P(E_1|H \wedge E_2) = P(E_1|H)$$
$$P(E_2|H \wedge E_1) = P(E_2|H)$$

In general we can write:

$$P(H|E_1,E_{2\ldots.}\ E_k) = \alpha\ P(H) \prod_{k=1}^{n} P(E_k|H)$$

Ideal Bayesian Decision-Making System

Bayesian Inference Network

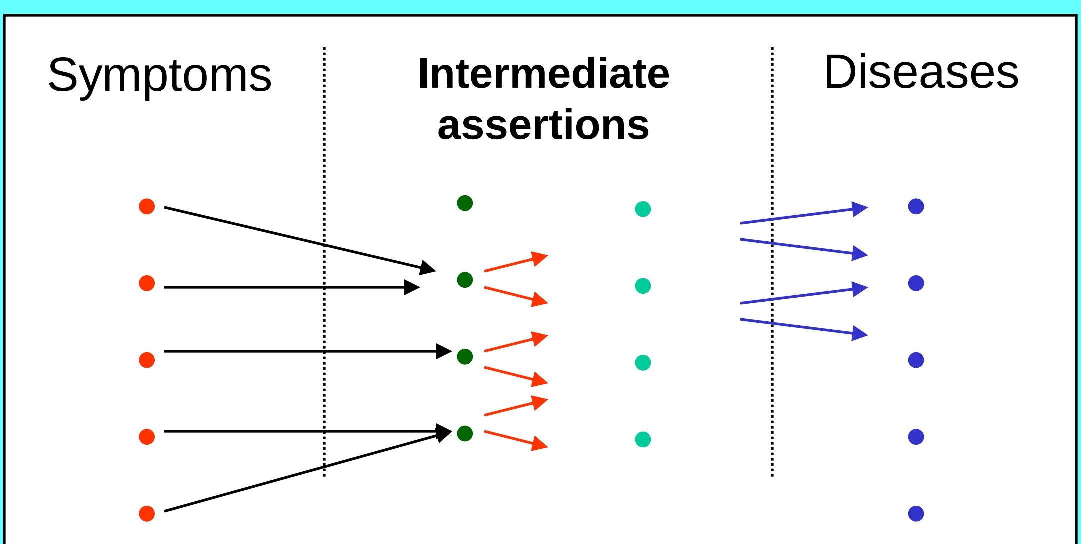


Realistic Bayesian Decision-Making System

Bayesian Inference Network

- Ideal Bayesian system unattainable – incomplete knowledge;
- Use heuristic modelling tools to ‘fill the gaps’ in knowledge base.

Simplistically……..



Realistic Bayesian Decision-Making System

Steps in design of Inference Network

1. **Input evidence**
2. **Decision alternatives**
3. **Intermediate assertions**
4. **Inference limits**
5. **Tune probabilities / inference function**

Realistic Bayesian Decision-Making System

Example: The Car Doctor -1

Symptoms :

- S1 clanking engine noise
- S2 car low on pick-up
- S3 poor starting
- S4 parts difficult to obtain

What is the truth of ……..

- C1 the repairs costs over £1000

  - difficult to infer C1 directly from S1….S4

3 'First level' hypotheses :

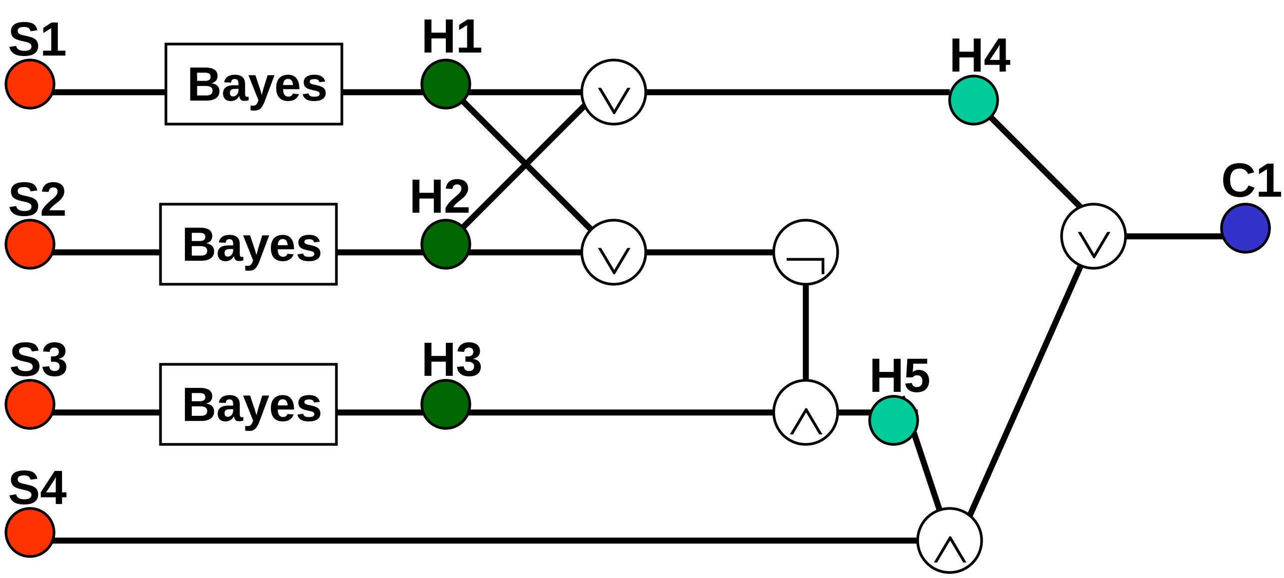
- H1 broken con rod
- H2 worn camshaft
- H3 car is out of tune

Realistic Bayesian Decision-Making System

Example: The Car Doctor -2

2 'Second level' hypotheses :

- H4 need to replace engine
- H5 engine needs retuning



In designing this system we would first ask our human expert a large set of questions, so that as the AI expert, we can better understand his/her decision-making process. We might ask schematically what evidence is used, in our simple case this is represented as S1 - S4, and what inferences can be made from this evidence, H1-H5, before linking this to a prognosis or outcome, C1.

Realistic Bayesian Decision-Making System

Example: The Car Doctor - 3

A	B	¬A	A∧B	A∨B	A→B	A⊕B
F	F	T	F	F	T	F
F	T	T	F	T	T	T
T	F	F	F	T	F	T
T	T	F	T	T	T	F

Recall from the Logic lectures, that when we have a set of propositions, lets call them A, and B, then we can link these together into sentences as conjuncts or disjuncts using logical connectives such as AND, OR and IMPLIES etc. YOU SHOULD KNOW THESE TRUTH TABLES!

In our example here, we use 'symptoms' which represent our observation propositions, so that these are either present or not present, I.e. either asserted or not asserted.

We will shortly use Bayes' Theorem to move these to posterior probabilities. In other words, given a set of priors (frequency of a given symptom), we can take the known probability, say, P(S|H1) and from this compute P(H1|S), which is the probability of finding a broken con rod given evidence of some combination of symptoms such as poor starting, noisy engine etc.

However, we must first fill in some blanks on the inference network…..

Realistic Bayesian Decision-Making System

Example: The Car Doctor - 4

Our next problem is to try to deal with intermediate assertions. These represent states which are intermediate between diagnosing the engine fault, and finding the prognosis or outcome (in our case that the engine repair cost will be >£1000).

There may be no clear way in which our expert makes these inferences. However we need to somehow model the experts rather 'blurry' or 'fuzzy' decision making process. We therefore resort to using fuzzy inference rules. These have been shown, empirically, to imitate the way in which humans make inferences given rather imprecise evidence.

Fuzzy inference rules, in general, take a set of probabilities, n, as arguments and maps them onto a single probability value, using some function,f. The appropriate choice of function depends on the particular application.

Formally, we can write $\mathbf{f:[0,1]^n} \longrightarrow [0,1]$.

There are 2 common inference rules which have been used in the past with some success in different applications. These two rules are sometimes referred to as the possibilistic fuzzy inference rule and the probabilistic fuzzy inference rule. We'll add these onto the previous table now….

Realistic Bayesian Decision-Making System

Example: The Car Doctor - 5

A	B	¬A	A∧B	A∨B	A→B	A⊕B
F	F	T	F	F	T	F
F	T	T	F	T	T	T
T	F	F	F	T	F	T
T	T	F	T	T	T	F
a **Possibilistic**	b	1-a	min (a,b)	max (a,b)	max (1-a,b)	xor (a,b)
a **Probabilistic**	b	1-a	ab	a+b-ab	1-a+ab	Xor(a,b)

A and B considered dependent
A and B considered independent

XOR means - 'whenever A or B occurs, but never both'. We can represent this as a fuzzy inference rule in 2 possible ways:

Possibilistic: xor(a,b) = max[min(a,1-b), min(1-a,b)]

Probabilistic: $Xor(a,b) = a + b - 2ab + a^2b + ab^2 – a^2b^2$

Realistic Bayesian Decision-Making System

Example: The Car Doctor -6

S1	S2	S3	P(S|H1) P(H1)=0.0001	P(S|H2) P(H2)=0.0002	P(S|H3) P(H3)=0.1	P(S)
F	F	F	0.001	0.200	0.2	0.4405
F	F	T	0.003	0.100	0.2	0.250
F	T	F	0.006	0.100	0.2	0.109
F	T	T	0.15	0.100	0.396	0.200
T	F	F	0.04	0.125	0.001	0.0001
T	F	T	0.06	0.125	0.001	0.0001
T	T	F	0.11	0.125	0.001	0.0001
T	T	T	0.63	0.125	0.001	0.0002

Following on from page 3 of the problem, we now take the basic set of propositions, and using data acquired from, say, national car service centres, and/or our human experts opinions, we can produce a set of priors, and conditional probabilities. Note each column sums to 1.

Next we'll systematically apply Bayes' rule to compute the posterior probabilities, and use for this application, our possibilistic inference rule to make intermediate inferences. We'll then show that we have all the information needed to compute our conclusion or outcome, C1.

Realistic Bayesian Decision-Making System

Example: The Car Doctor -7

C1: What is the truth of the statement :

The estimated repair costs will be over >£1000

From previously shown inference net -

(a) $H_4 = H_1 \vee H_2$ **(b)** $H_5 = \neg(H_1 \vee H_2) \wedge H_3$

For our problem we will consider the particular case that all symptoms are present. i.e. S_{123}=T.

Now evaluate $P(H_4|S)$, using (a) ……

$$P(H_4 \mid S_{123}) = \max\,[P(H_1|S_{123}), P(H_2|S_{123})]$$

e.g. first use Bayes' rule to find $P(H_1| S) = \dfrac{P(S \mid (H_1)\, P(H_1))}{P(S)}$

$$\frac{0.63 \;\; 0.0001}{0.0002}$$

$$\boxed{\mathbf{P(H_1| S) = \;\; 0.315}}$$

Similarly use Bayes Rule to compute $P(H_2|S_{123})$

Realistic Bayesian Decision-Making System

Example: The Car Doctor -8

similarly $\boxed{\mathbf{P(H_2|S) = 0.125}}$

Now **$P(H_4 \mid S)$** $= \max [P(H_1| S), P(H_2| S)]$

$= \max [0.315, 0.125]$

$\underline{= 0.315}$

Using (b) on last page and calculating $P(H_5|S)$ results in

$P(H_5|S) = 0.5$ *Exercise: try calculating this yourself.*

Now **$P(C_1 \mid S)$** $= H_4 \vee [H_5 \wedge S_4]$

$= \max [P(H_4|S), \min(P(H_5|S),v)]$

$= \max [0.315, \min (0.5, 1)]$ *- if S_4 is true, $v = 1$*

$\underline{= 0.5}$

- so in 50% of cases <u>when all symptoms S1-S4</u> are present, then the cost of repair will be >£1000

Odds & Bayes Rule - 1

Experts often give unreliable probabilities – much better to reformulate problem in terms of odds

Certain evidence $\quad E \xrightarrow{\lambda\ ,\ \lambda'} H \quad$ Probabilistic rule

$$P(H\,|E) = \frac{P(E|H)\,P(H)}{P(E)} \qquad \ldots.\ 1$$

$$P(\neg H\,|E) = \frac{P(E|\neg H)\,P(\neg H)}{P(E)} \qquad \ldots.\ 2$$

Define ‘Odds’ as:

$$O(x) = \frac{P(x)}{1 - P(x)} \qquad \ldots.\ 3$$

So, if we divide eq(1) by eq (2) we can write:

likelihood ratio

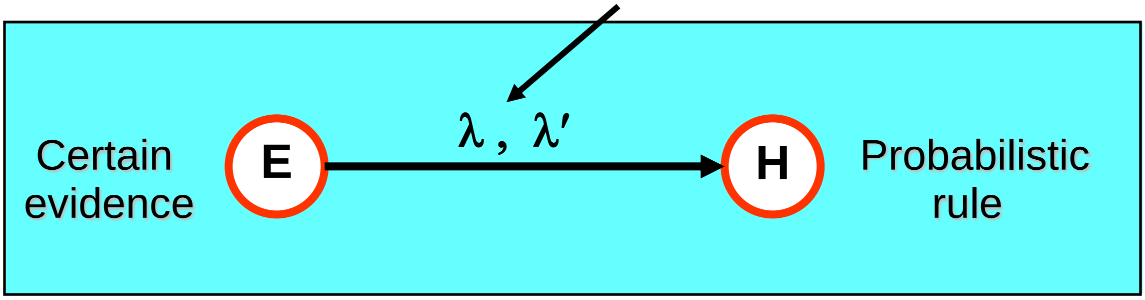
$$\text{Odds, } O(H|E) = \boxed{\frac{P(E|H)}{P(E|\neg H)}}\ O(H) = \lambda\ O(H)$$

prior odds on H

Odds & Bayes Rule - 2

Likelihood ratio, sufficiency and necessity

Indicates how presence or absence of evidence influences odds on hypothesis



$\lambda > 1$ → presence of evidence reinforces belief in H

$\lambda \cong 0$ → reduces belief in H

- sufficiency coefficient

However, if E is false or known not to be present, then

$O(H \mid \neg E) = \lambda' \, O(H)$, where - λ' necessity coefficient

$$\lambda' = \frac{P(\neg E \mid H)}{P(\neg E \mid \neg H)} = \frac{1 - P(E \mid H)}{1 - P(E \mid \neg H)}$$

Summary / Conclusions

Probabilistic vs Logical reasoning

- **If I have toothache, then I have a cavity;**
- **If I have toothache, then there's a 95% probability its a cavity.**

Bayes' rule used for diagnosis because….

- **P(H|E) is what experts (e.g. doctors) do - difficult;**
- **P(E|H) is often easier to determine, as are P(E), P(H)**

- **We rarely have all the information needed for just using Bayes' Rule. We have to make modelling decisions to represent the human decision making process.**
- **Use heuristics to assist problem. One way to achieve this is to use Fuzzy Inference Rules, to infer some 'degree of truth' about a proposition al level k+1, from substantiated hypotheses made at level k.**
- **Bayesian networks/classifiers used for more complex problems.**